

# ON THE UNION OF INTERSECTING FAMILIES

DAVID ELLIS AND NOAM LIFSHITZ

**ABSTRACT.** We prove that for any integer  $r \geq 2$ , if  $X$  is an  $n$ -element set, and  $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2 \cup \dots \cup \mathcal{F}_r$ , where each  $\mathcal{F}_i$  is an intersecting family of  $k$ -element subsets of  $X$ , then  $|\mathcal{F}| \leq \binom{n}{k} - \binom{n-r}{k}$ , provided  $n \geq 2k + Ck^{2/3}$ , where  $C$  is a constant depending upon  $r$  alone. Equality holds if and only if  $\mathcal{F} = \{S \subset X : |S| = k, S \cap R \neq \emptyset\}$  for some  $R \subset X$  with  $|R| = r$ . In the case  $r = 2$ , this improves a result of Frankl and Füredi.

## 1. INTRODUCTION

Let  $[n] := \{1, 2, \dots, n\}$ , and let  $\binom{[n]}{k} := \{S \subset [n] : |S| = k\}$ . If  $X$  is a set, we let  $\mathcal{P}(X)$  denote the power-set of  $X$ . A family  $\mathcal{F} \subset \mathcal{P}([n])$  is said to be *1-intersecting* (or just *intersecting*) if for any  $A, B \in \mathcal{F}$ ,  $A \cap B \neq \emptyset$ .

One of the best-known theorems in extremal combinatorics is the Erdős-Ko-Rado theorem [7], which bounds the size of an intersecting subfamily of  $\binom{[n]}{k}$ .

**Theorem 1** (Erdős-Ko-Rado, 1961). *Let  $k, n \in \mathbb{N}$  with  $k < n/2$ . If  $\mathcal{F} \subset \binom{[n]}{k}$  is intersecting, then  $|\mathcal{F}| \leq \binom{n-1}{k-1}$ . Equality holds only if  $\mathcal{F} = \{S \in \binom{[n]}{k} : j \in S\}$  for some  $j \in [n]$ .*

In [11], Frankl and Füredi consider the problem, first raised by Erdős [6], of determining the maximum possible size of a union of  $r$  1-intersecting subfamilies of  $\binom{[n]}{k}$ , for each triple of integers  $(n, k, r)$ . They prove the following.

**Theorem 2** (Frankl, Füredi, 1986). *If  $\mathcal{F} \subset \binom{[n]}{k}$  is a union of two intersecting families, and  $n > \frac{1}{2}(3 + \sqrt{5})k \approx 2.62k$ , then  $|\mathcal{F}| \leq \binom{n}{k} - \binom{n-2}{k}$ . Equality holds only if  $\mathcal{F} = \{S \in \binom{[n]}{k} : S \cap \{i, j\} \neq \emptyset\}$ , for some distinct  $i, j \in [n]$ .*

They give an example which shows that the upper bound in Theorem 2 does not hold provided if  $n_0 \leq n \leq 2k + c_0\sqrt{k}$ , where  $n_0, c_0 > 0$  are absolute constants with  $n_0$  sufficiently large and  $c_0$  sufficiently small; this disproved a conjecture of Erdős in [6].

In this paper, we prove the following strengthening and generalisation of Theorem 2.

**Theorem 3.** *For each integer  $r \geq 2$ , there exists a constant  $C = C(r) > 0$  such that the following holds. Let  $n \geq 2k + Ck^{2/3}$ , and let  $\mathcal{F} \subset \binom{[n]}{k}$  be a union of at most  $r$  1-intersecting families. Then  $|\mathcal{F}| \leq \binom{n}{k} - \binom{n-r}{k}$ , and equality holds only if  $\mathcal{F} = \{S \in \binom{[n]}{k} : S \cap R \neq \emptyset\}$  for some  $R \in \binom{[n]}{r}$ .*

We note that even in the case  $r = 2$ , the conclusion of Theorem 3 was previously known to hold only for  $n - 2k = \Omega(k)$ ; we prove it for  $n - 2k = o(k)$ , for any

fixed  $r \geq 2$ , though the correct rate of growth of the  $o(k)$  term remains open. We conjecture that the conclusion of Theorem 3 holds for  $n \geq 2k + c\sqrt{k}$  for  $c = c(r)$  sufficiently large; this would be best-possible up to the value of  $c$ , as evidenced by the aforementioned construction of Frankl and Füredi. It would be of great interest to determine the extremal families for every triple of integers  $(n, k, r)$ .

We remark that it is well-known that if  $\mathcal{F} \subset \mathcal{P}([n])$  is a union of at most  $r$  1-intersecting subfamilies of  $\mathcal{P}([n])$ , then  $|\mathcal{F}| \leq 2^n - 2^{n-r}$ ; this is an easy consequence of the FKG inequality (see Lemma 19), and is sharp, as evidenced by taking  $\mathcal{F} = \cup_{i=1}^r \{S \subset [n] : i \in S\}$ . In fact, we will use this bound in our proof of Theorem 3.

We remark also that the problem considered here is closely related to the well-known Erdős matching conjecture. Recall that the *matching number*  $m(\mathcal{F})$  of a family  $\mathcal{F} \subset \mathcal{P}([n])$  is defined to be the maximum integer  $s$  such that  $\mathcal{F}$  contains  $s$  pairwise disjoint sets. The 1965 *Erdős matching conjecture* [5] asserts that if  $n, k, s \in \mathbb{N}$  with  $n \geq (s+1)k$  and  $\mathcal{F} \subset \binom{[n]}{k}$  with  $m(\mathcal{F}) \leq s$ , then

$$|\mathcal{F}| \leq \max \left\{ \binom{n}{k} - \binom{n-s}{k}, \binom{k(s+1)-1}{k} \right\}.$$

This conjecture remains open. Erdős himself proved the conjecture for all  $n$  sufficiently large, i.e. for  $n \geq n_0(k, s)$ . The bound on  $n_0(k, s)$  was lowered in several works: Bollobás, Daykin and Erdős [2] showed that  $n_0(k, s) \leq 2sk^3$ ; Huang, Loh and Sudakov [13] showed that  $n_0(k, s) \leq 3sk^2$ , and Frankl and Füredi (unpublished) showed that  $n_0(k, s) = O(k s^2)$ . Perhaps the most significant result to date is the following theorem of Frankl [10]:

**Theorem 4** (Frankl, 2013). *Let  $n, k, s \in \mathbb{N}$  such that  $n \geq (2s+1)k - s$ , and let  $\mathcal{F} \subset \binom{[n]}{k}$  such that  $m(\mathcal{F}) \leq s$ . Then  $|\mathcal{F}| \leq \binom{n}{k} - \binom{n-s}{k}$ . Equality holds if and only if there exists  $S \in \binom{[n]}{s}$  such that  $\mathcal{F} = \{F \in \binom{[n]}{k} : F \cap S \neq \emptyset\}$ .*

Clearly, if  $\mathcal{F} \subset \binom{[n]}{k}$  is a union of at most  $r$  1-intersecting families, then  $m(\mathcal{F}) \leq r$ , so Theorem 4 implies the conclusion of Theorem 3 under the (stronger) condition  $n \geq (2r+1)k - r$ .

**Our proof techniques.** Our main tool is the following ‘stability’ version of Theorem 3.

**Theorem 5.** *There exists an absolute constant  $C_0$  such that the following holds. Let  $r, k \in \mathbb{N}$ , let  $s \geq C_0 \sqrt{\log k}$ , let  $t \in \mathbb{N}$  with  $t \geq s^2$ , let  $n \geq 2k + s\sqrt{k}$ , and let  $\mathcal{F} \subset \binom{[n]}{k}$  be a family satisfying  $\mu_{\frac{1}{2}}(\mathcal{F}^\uparrow) \leq 1 - 2^{-r}$  and  $|\mathcal{F}| \geq \binom{n}{k} - \binom{n-r}{k} - \binom{n-r-t}{k-1}$ . Then there exists  $R \in \binom{[n]}{r}$  such that  $|\{S \in \mathcal{F} : S \cap R = \emptyset\}| \leq 2^r \exp(-\Theta(s^2)) \binom{n-r}{k}$ .*

Roughly speaking, our strategy for proving Theorem 5 is as follows. Instead of working with the uniform measure on  $\binom{[n]}{k}$ , we take the upward monotone closure  $\mathcal{F}^\uparrow$  of our family  $\mathcal{F}$ , and we work with the  $p$ -biased measure on  $\mathcal{P}([n])$ , where  $p \approx k/n$  (see Section 2 for the definition of the  $p$ -biased measure); it is well-known that  $\mu_p(\mathcal{F}^\uparrow)$  approximately bounds  $|\mathcal{F}|/\binom{n}{k}$  from above, for an appropriate choice of  $p$ . More precisely, we choose  $p$  to be slightly larger than  $k/n$ , and use the lower bound on  $|\mathcal{F}|$  to show that  $\mu_p(\mathcal{F}^\uparrow) \approx 1 - (1-p)^r$ . Combined with the fact that  $\mu_{1/2}(\mathcal{F}^\uparrow) \leq 1 - 2^{-r}$ , this implies an upper bound on the derivative of the function  $q \mapsto \mu_q(\mathcal{F}^\uparrow)$ , at some  $q \in (p, 1/2)$ . But by Russo’s Lemma, this derivative is precisely  $I^q[\mathcal{F}^\uparrow]$ , the *influence* of  $\mathcal{F}^\uparrow$  with respect to the  $q$ -biased measure; we

deduce that  $I^q[\mathcal{F}^\uparrow]$  is close to its minimum possible value. We then use a recent structure theorem for families with small influence (proved in [4]) to deduce that  $\mathcal{F}^\uparrow$  must be close (with respect to the  $q$ -biased measure) to a family of the form  $\{S \subset [n] : S \cap R \neq \emptyset\}$ , for some  $R \in \binom{[n]}{r}$ . Finally, we deduce from this that  $\mathcal{F}$  is almost contained in a family of the form  $\{S \in \binom{[n]}{k} : S \cap R \neq \emptyset\}$ . Note that a similar strategy was used to obtain the stability results in [3].

We deduce Theorem 3 from Theorem 5, together with an additional ‘bootstraping’ argument.

## 2. DEFINITIONS, NOTATION AND TOOLS

**Definitions and notation.** In this paper, all logs are to base 2. A *dictatorship* is a family of the form  $\{S \subset [n] : j \in S\}$  or  $\{S \in \binom{[n]}{k} : j \in S\}$  for some  $j \in [n]$ . For  $j \in [n]$ , we write  $\mathcal{D}_j := \{S \in \binom{[n]}{k} : j \in S\}$  for the corresponding dictatorship. If  $R \subset [n]$ , we write  $\mathcal{S}_R := \{S \subset [n] : R \subset S\}$ , and we write  $\text{OR}_R := \{S \subset [n] : S \cap R \neq \emptyset\}$ .

A family  $\mathcal{F} \subset \mathcal{P}([n])$  is said to be *increasing* (or an *up-set*) if it is closed under taking supersets, i.e. whenever  $A \subset B$  and  $A \in \mathcal{F}$ , we have  $B \in \mathcal{F}$ ; it is said to be *decreasing* (or a *down-set*) if it is closed under taking subsets.

If  $\mathcal{F} \subset \mathcal{P}([n])$  and  $l \in [n]$ , we write  $\mathcal{F}^{(l)} := \{F \in \mathcal{F} : |F| = l\}$ . Hence, for example,

$$(\text{OR}_{[r]})^{(k)} = \{S \in \binom{[n]}{k} : S \cap R \neq \emptyset\}.$$

If  $\mathcal{F} \subset \mathcal{P}([n])$ , we define the *dual* family  $\mathcal{F}^*$  by  $\mathcal{F}^* = \{[n] \setminus A : A \notin \mathcal{F}\}$ . We denote by  $\mathcal{F}^\uparrow$  the up-closure of  $\mathcal{F}$ , i.e. the minimal increasing subfamily of  $\mathcal{P}([n])$  which contains  $\mathcal{F}$ .

If  $\mathcal{F} \subset \mathcal{P}([n])$  and  $C \subset B \subset [n]$ , we define  $\mathcal{F}_B^C := \{S \in \mathcal{P}([n] \setminus B) : S \cup C \in \mathcal{F}\}$ .

A family  $\mathcal{F} \subset \mathcal{P}([n])$  is said to be a *subcube* if  $\mathcal{F} = \{S \subset [n] : S \cap B = C\}$ , for some  $C \subset B \subset [n]$ .

We say a pair of families  $\mathcal{A}, \mathcal{B} \subset \mathcal{P}([n])$  are *cross-intersecting* if  $A \cap B \neq \emptyset$  for any  $A \in \mathcal{A}$  and any  $B \in \mathcal{B}$ .

If  $\mathcal{A} \subset \mathcal{P}([n])$ , we write  $1_{\mathcal{A}}$  for the *indicator function* of  $\mathcal{A}$ , i.e., the Boolean function

$$1_{\mathcal{A}} : \mathcal{P}([n]) \rightarrow \{0, 1\}; \quad 1_{\mathcal{A}}(x) = \begin{cases} 1 & \text{if } x \in \mathcal{A}; \\ 0 & \text{if } x \notin \mathcal{A}. \end{cases}$$

By identifying  $\{0, 1\}^n$  with  $\mathcal{P}([n])$  in the usual way (identifying a vector  $x \in \{0, 1\}^n$  with the set  $\{i : x_i = 1\} \subset [n]$ ), we may identify Boolean functions on  $\{0, 1\}^n$  with Boolean functions on  $\mathcal{P}([n])$ , and therefore with subfamilies of  $\mathcal{P}([n])$ . We will sometimes write Boolean functions on  $\{0, 1\}^n$  using the AND ( $\wedge$ ) and OR ( $\vee$ ) operators. Hence, for example,

$$f : \{0, 1\}^n \rightarrow \{0, 1\}; \quad f(x_1, \dots, x_n) \mapsto x_1 \vee (x_2 \wedge x_3)$$

corresponds to the subfamily  $\{S \subset [n] : 1 \in S \text{ or } \{2, 3\} \subset S\} \subset \mathcal{P}([n])$ .

For  $p \in [0, 1]$ , the *p-biased measure* on  $\mathcal{P}([n])$  is defined by

$$\mu_p(S) = p^{|S|}(1-p)^{n-|S|} \quad \forall S \subset [n].$$

In other words, we choose a random set by including each  $j \in [n]$  independently with probability  $p$ . For  $\mathcal{F} \subset \mathcal{P}([n])$ , we define  $\mu_p(\mathcal{F}) = \sum_{S \in \mathcal{F}} \mu_p(S)$ .

We remark that if  $C \subset B \subset [n]$ , then  $\mu_p(\mathcal{F}_B^C)$  refers to the  $p$ -biased measure on  $\mathcal{P}([n] \setminus B)$ , not on  $\mathcal{P}([n])$ , since we regard  $\mathcal{F}_B^C$  as a subset of  $\mathcal{P}([n] \setminus B)$ .

If  $f : \mathcal{P}([n]) \rightarrow \{0, 1\}$  is a Boolean function, we define the *influence of  $f$  in direction  $i$*  (with respect to  $\mu_p$ ) by

$$\text{Inf}_i^p[f] := \mu_p(\{S \subset [n] : f(S) \neq f(S \Delta \{i\})\}).$$

We define the *total influence* of  $f$  (w.r.t.  $\mu_p$ ) by  $I^p[f] := \sum_{i=1}^n \text{Inf}_i^p[f]$ .

Similarly, if  $\mathcal{A} \subset \mathcal{P}([n])$ , we define the *influence of  $\mathcal{A}$  in direction  $i$*  (w.r.t.  $\mu_p$ ) by  $\text{Inf}_i^p[\mathcal{A}] := \text{Inf}_i^p[1_{\mathcal{A}}]$ , and we define *total influence* of  $\mathcal{A}$  (w.r.t.  $\mu_p$ ) by  $I^p[\mathcal{A}] := I^p[1_{\mathcal{A}}]$ .

**Tools.** We will use the following ‘biased version’ of the Erdős-Ko-Rado theorem, first obtained by Ahlswede and Katona [1] in 1977.

**Theorem 6.** *Let  $0 < p \leq 1/2$ . Let  $\mathcal{F} \subset \mathcal{P}([n])$  be an intersecting family. Then  $\mu_p(\mathcal{F}) \leq p$ . If  $p < 1/2$ , then equality holds if and only if  $\mathcal{F} = \{S \subset [n] : j \in S\}$  for some  $j \in [n]$ .*

We will use the following special case of the well-known inequality of Harris [12] (which is itself a special case of the FKG inequality [8]).

**Lemma 7** (Harris’ inequality). *Let  $0 < p < 1$ . Then for any increasing sets  $\mathcal{A}, \mathcal{B} \subset \mathcal{P}([n])$ ,  $\mu_p(\mathcal{A} \cap \mathcal{B}) \geq \mu_p(\mathcal{A})\mu_p(\mathcal{B})$ . The same inequality holds if  $\mathcal{A}$  and  $\mathcal{B}$  are decreasing.*

By repeatedly applying Lemma 7, one immediately obtains the following well-known corollary.

**Corollary 8.** *Let  $r \in \mathbb{N}$ , let  $0 < p < 1$ , and suppose  $\mathcal{A}_1, \dots, \mathcal{A}_r \subset \mathcal{P}([n])$  are increasing. Then*

$$\mu_p(\mathcal{A}_1 \cap \dots \cap \mathcal{A}_r) \geq \prod_{i=1}^r \mu_p(\mathcal{A}_i).$$

*The same inequality holds if  $\mathcal{A}_1, \dots, \mathcal{A}_r$  are decreasing.*

The following ‘biased isoperimetric inequality’ is well-known; it appears for example in [14].

**Theorem 9.** *If  $0 < p < 1$  and  $\mathcal{A} \subset \mathcal{P}([n])$  is increasing, then*

$$(2.1) \quad pI_p[\mathcal{A}] \geq \mu_p(\mathcal{A}) \log_p(\mu_p(\mathcal{A})).$$

We will need the following ‘stability’ version of Theorem 9, proved in [4].

**Theorem 10.** *For each  $\eta > 0$ , there exist  $C_1 = C_1(\eta)$ ,  $c_0 = c_0(\eta) > 0$  such that the following holds. Let  $0 < p \leq 1 - \eta$ , and let  $0 \leq \epsilon \leq c_0 / \ln(1/p)$ . Let  $\mathcal{A} \subset \mathcal{P}([n])$  be an increasing family such that*

$$pI^p[\mathcal{A}] \leq \mu_p(\mathcal{A}) (\log_p(\mu_p(\mathcal{A})) + \epsilon).$$

*Then there exists an increasing subcube  $\mathcal{C} \subset \mathcal{P}([n])$  such that*

$$\mu_p(\mathcal{A} \Delta \mathcal{C}) \leq \frac{C_1 \epsilon \ln(1/p)}{\ln\left(\frac{1}{\epsilon \ln(1/p)}\right)} \mu_p(\mathcal{A}).$$

We will need the well-known lemma of Russo [15], which relates the derivative of the function  $p \mapsto \mu_p(A)$  to the total influence  $I^p(A)$ , where  $A \subset \{0, 1\}^n$  is increasing.

**Lemma 11** (Russo's lemma). *Let  $\mathcal{A} \subset \mathcal{P}([n])$  be increasing, and let  $0 < p_0 < 1$ . Then*

$$\left. \frac{d\mu_p(\mathcal{A})}{dp} \right|_{p=p_0} = I^{p_0}[\mathcal{A}].$$

We need the following lemma from [3], which follows from Russo's lemma and Theorem 9.

**Lemma 12.** *If  $\mathcal{A} \subset \mathcal{P}([n])$  is increasing, then the function  $p \mapsto \log_p(\mu_p(\mathcal{A}))$  is monotone non-increasing on  $(0, 1)$ .*

We will also need the following Chernoff bound.

**Lemma 13.** *Let  $n \in \mathbb{N}$ , let  $0 < \delta, p < 1$  and let  $X \sim \text{Bin}(n, p)$ . Then*

$$(2.2) \quad \Pr[X \leq (1 - \delta)np] < e^{-\delta^2 np/2}.$$

The following lemma (combined with the Chernoff bound (2.2)) will allow us to bound  $|\mathcal{G}|/\binom{n}{k}$  from above in terms of  $\mu_p(\mathcal{G}^\dagger)$ , where  $\mathcal{G} \subset \binom{[n]}{k}$  and  $p$  is slightly larger than  $k/n$ .

**Lemma 14.** *Let  $k, n \in \mathbb{N}$ , let  $0 < \epsilon, p < 1$  and let  $\mathcal{G} \subset \binom{[n]}{k}$  be a family with  $|\mathcal{G}| = \alpha \binom{n}{k}$ . Then*

$$\mu_p(\mathcal{G}^\dagger) \geq \alpha \Pr[\text{Bin}(n, p) \geq k].$$

*Proof.* For each  $l \geq k$ , the local LYM inequality implies that  $|(\mathcal{G}^\dagger)^{(l)}|/\binom{n}{l} \geq |\mathcal{G}|/\binom{n}{k} = \alpha$ . Hence,

$$\begin{aligned} \mu_p(\mathcal{G}^\dagger) &\geq \sum_{l=k}^n p^l (1-p)^{n-l} \alpha \binom{n}{l} \\ &= \alpha \Pr[\text{Bin}(n, p) \geq k], \end{aligned}$$

as required.  $\square$

Finally, we need the following immediate consequence of a lemma of Hilton (see [9]).

**Lemma 15.** *Let  $n, k, l, t \in \mathbb{N}$  with  $k + l \leq n$ . Let  $\mathcal{A} \subset \binom{[n]}{k}$ ,  $\mathcal{B} \subset \binom{[n]}{l}$  be cross-intersecting families. If  $|\mathcal{A}| \geq \binom{n}{k} - \binom{n-t}{k}$ , then  $|\mathcal{B}| \leq \binom{n-t}{l-t}$ .*

### 3. PROOFS OF THE MAIN RESULTS

Our first aim is to prove Theorem 5; for this, we need some preliminary lemmas.

**Lemma 16.** *Let  $s > 0$  and let  $t \in \mathbb{N}$  with  $t \geq s^2$ . Let  $n, k \in \mathbb{N}$  with  $n \geq 2k + s\sqrt{k}$ , and let  $p = \frac{k/n+0.5}{2}$ . If  $\mathcal{F} \subset \binom{[n]}{k}$  with  $|\mathcal{F}| \geq \binom{n}{k} - \binom{n-r}{k} - \binom{n-r-t}{k-1}$ , then*

$$\mu_p(\mathcal{F}^\dagger) \geq 1 - (1-p)^r - \exp(-\Theta(s^2)).$$

*Proof.* The Kruskal-Katona Theorem implies that

$$\begin{aligned} (\mathcal{F}^\dagger)^{(l)} &\geq \binom{n}{l} - \binom{n-r}{l} - \binom{n-r-t}{l-1} \\ &= | (x_1 \vee x_2 \vee \dots \vee x_{r-1} \vee (x_r \wedge (x_{r+1} \vee x_{r+2} \vee \dots \vee x_{r+t})))^{(l)} | \end{aligned}$$

for any  $l \geq k$ . It follows that

$$\begin{aligned} \mu_p(\mathcal{F}^\dagger) &\geq \mu_p(x_1 \vee x_2 \vee \dots \vee x_{r-1} \vee (x_r \wedge (x_{r+1} \vee x_{r+2} \vee \dots \vee x_{r+t}))) \\ &\quad - \Pr[\text{Bin}(n, p) < k] \\ &= 1 - (1-p)^r - p(1-p)^{r+t-1} - \Pr[\text{Bin}(n, p) < k]. \end{aligned}$$

The Chernoff bound (2.2), together with our condition on  $t$ , completes the proof.  $\square$

**Lemma 17.** *Let  $r, n \in \mathbb{N}$ , let  $0 < p < 1/2$  and let  $0 < \eta < 1$ . If  $\mathcal{A} \subset \mathcal{P}([n])$  is increasing with  $\mu_{1/2}(\mathcal{A}) \leq 1 - 2^{-r}$  and*

$$\mu_p(\mathcal{A}) \geq 1 - (1-p)^r - \eta,$$

*then there exists  $p' \in (p, \frac{1}{2})$  such that*

$$I^{p'}[\mathcal{A}] \leq I^{p'}[x_1 \vee \dots \vee x_r] + \frac{\eta}{0.5-p}.$$

*Proof.* By Russo's lemma (Lemma 17 above), we have

$$\int_p^{0.5} I^q[\mathcal{A}] dq = \mu_{\frac{1}{2}}(\mathcal{A}) - \mu_p(\mathcal{A}) \leq 1 - 2^{-r} - (1 - (1-p)^r) + \eta.$$

Hence,

$$\int_p^{0.5} (I^q[\mathcal{A}] - I^q[x_1 \vee \dots \vee x_r]) dq \leq \eta.$$

This implies that for some  $p' \in (p, 0.5)$  we have

$$I^{p'}[\mathcal{A}] - I^{p'}[x_1 \vee \dots \vee x_r] \leq \frac{\eta}{0.5-p},$$

as required.  $\square$

**Lemma 18.** *There exist absolute constants  $\delta_0, \epsilon_0, C_2 > 0$  such that the following holds. Let  $0 \leq \delta < \delta_0$ ,  $0 \leq \epsilon < \epsilon_0$  and  $1/4 \leq p < p' < 1/2$ . If  $\mathcal{A} \subset \mathcal{P}([n])$  is increasing with  $\mu_{1/2}(\mathcal{A}) \leq 1 - 2^{-r}$ ,  $\mu_p(\mathcal{A}) \geq 1 - (1-p)^r(1+\delta)$  and*

$$I^{p'}[\mathcal{A}] - I^{p'}[x_1 \vee \dots \vee x_r] \leq \epsilon(1-p')^r,$$

*then there exists  $R \in \binom{[n]}{r}$  such that*

$$\mu_{p'}(\mathcal{A}_R^\emptyset) \leq C_2(\epsilon + \delta).$$

*Proof.* Note that for any family  $\mathcal{B} \subset \mathcal{P}([n])$ , we have  $I^{p'}[\mathcal{B}] = I^{1-p'}[\mathcal{B}^*]$ . Hence, by hypothesis, we have

$$I^{1-p'}[\mathcal{A}^*] - r(1-p')^{r-1} = I^{1-p'}[\mathcal{A}^*] - I^{1-p'}[x_1 \wedge \dots \wedge x_r] \leq \epsilon(1-p')^r.$$

Since  $\mathcal{A}^*$  is increasing and  $\mu_{1/2}(\mathcal{A}^*) = 1 - \mu_{1/2}(\mathcal{A}) \geq 2^{-r}$ , by Lemma 12 we have  $\mu_{1-p'}(\mathcal{A}^*) \geq (\mu_{1/2}(\mathcal{A}^*))^{\log_{1/2}(1-p')} = (1-p')^r$ . Similarly, since  $\mu_{1-p}(\mathcal{A}^*) = 1 - \mu_p(\mathcal{A}^*) \leq (1-p)^r(1+\delta)$ , we have

$$\mu_{1-p'}(\mathcal{A}^*) \leq (\mu_{1-p}(\mathcal{A}^*))^{\log_{1-p}(1-p')} \leq ((1-p)^r(1+\delta))^{\log_{1-p}(1-p')} \leq (1-p')^r(1+3\delta),$$

provided  $\delta_0$  is sufficiently small. Therefore,

$$\mu_{1-p'}(\mathcal{A}^*) \log_{1-p'}(\mu_{1-p'}(\mathcal{A}^*)) \geq (1-p')^r \log_{1-p'}((1-p')^r(1+3\delta)) \geq (1-5\delta)r(1-p')^r.$$

It follows that

$$(1-p')I^{1-p'}[\mathcal{A}^*] - \mu_{1-p'}(\mathcal{A}^*) \log_{1-p'}(\mu_{1-p'}(\mathcal{A}^*)) \leq (\epsilon + 5\delta)\mu_{1-p'}(\mathcal{A}^*).$$

Applying Theorem 10 (with  $\eta = 1/4$ , with  $1-p'$  in place of  $p$  and with  $\epsilon + 5\delta$  in place of  $\epsilon$ ) to the family  $\mathcal{A}^*$ , we see that there exists  $R \subset [n]$  such that

$$(3.1) \quad \mu_{1-p'}(\mathcal{A}^* \Delta \mathcal{S}_R) \leq C_2(\epsilon + \delta)(1-p')^r,$$

where  $C_2 > 0$  is an absolute constant, provided  $\epsilon_0, \delta_0$  are sufficiently small. We claim that  $|R| = r$ . Indeed, if  $|R| > r$ , then

$$\mu_{1-p'}(\mathcal{A}^* \Delta \mathcal{S}_R) \geq \mu_{1-p'}(\mathcal{A}^*) - \mu_{1-p'}(\mathcal{S}_R) \geq (1-p')^r - (1-p')^{r+1} = p'(1-p')^r,$$

contradicting (3.1) provided  $\epsilon_0, \delta_0$  are sufficiently small. Similarly, if  $|R| < r$ , then

$$\begin{aligned} \mu_{1-p'}(\mathcal{A}^* \Delta \mathcal{S}_R) &\geq \mu_{1-p'}(\mathcal{S}_R) - \mu_{1-p'}(\mathcal{A}^*) \\ &\geq (1-p')^{r-1} - (1+3\delta)(1-p')^r \\ &= (1-p')^{r-1}(p' - 3(1-p')\delta), \end{aligned}$$

again contradicting (3.1) provided  $\epsilon_0, \delta_0$  are sufficiently small. This proves the claim. It follows that

$$\begin{aligned} \mu_{p'}(\mathcal{A}_R^\emptyset) &= (1-p')^{-r} \mu_{p'}(\mathcal{A} \setminus \text{OR}_R) \\ &\leq (1-p')^{-r} \mu_{p'}(\mathcal{A} \Delta \text{OR}_R) \\ &= (1-p')^{-r} \mu_{1-p'}(\mathcal{A}^* \Delta \mathcal{S}_R) \\ &\leq C_2(\epsilon + \delta), \end{aligned}$$

as required.  $\square$

*Proof of Theorem 5.* Let  $n, k, r, s$  and  $t$  be as in the statement of the theorem, where  $C_0$  is to be chosen later. Let  $\mathcal{F} \subset \binom{[n]}{k}$  be a family satisfying  $\mu_{\frac{1}{2}}(\mathcal{F}^\uparrow) \leq 1 - 2^{-r}$  and  $|\mathcal{F}| \geq \binom{n}{k} - \binom{n-r}{k} - \binom{n-r-t}{k-1}$ .

Let  $p = \frac{k/n+0.5}{2}$ . By Lemma 16, we have

$$\mu_p(\mathcal{F}^\uparrow) \geq 1 - (1-p)^r - \exp(-\Theta(s^2)).$$

Applying Lemma 17 with  $\eta = \exp(-\Theta(s^2))$  and  $\mathcal{A} = \mathcal{F}^\uparrow$ , yields  $p' \in (p, \frac{1}{2})$  such that

$$I^{p'}[\mathcal{F}^\uparrow] \leq I^{p'}[x_1 \vee \dots \vee x_r] + \frac{\delta}{0.5-p}.$$

Provided  $C_0$  is sufficiently large, we may apply Lemma 18 with  $\delta = 2^r \exp(-\Theta(s^2))$  and

$$\epsilon = \frac{\exp(-\Theta(s^2))}{(0.5-p)(1-p')^r} \leq \frac{2^r \sqrt{k}}{s} \exp(-\Theta(s^2)) \leq 2^r \exp(-\Theta(s^2)),$$

yielding

$$\mu_{p'}((\mathcal{F}^\uparrow)_R^\mathcal{O}) \leq 2^r \exp(-\Theta(s^2))$$

for some  $p' \in (p, 1/2)$  and some  $R \in \binom{[n]}{r}$ .

Using the fact that  $p' - k/n > p - k/n = \Omega\left(\frac{s}{\sqrt{k}}\right)$ , and applying Lemma 14 and the Chernoff bound (2.2), we obtain

$$\frac{|\mathcal{F}_R^\mathcal{O}|}{\binom{n-r}{k}} = O(\mu_{p'}((\mathcal{F}^\uparrow)_R^\mathcal{O})) \leq 2^r \exp(-\Theta(s^2)),$$

completing the proof of Theorem 5.  $\square$

Before proving Theorem 3, we need some additional lemmas.

**The FKG bound.** We need the following well-known upper bound on the  $p$ -biased measure of the union of  $r$  1-intersecting subfamilies of  $\mathcal{P}([n])$ ; we provide a proof for completeness.

**Lemma 19.** *If  $\mathcal{F}_1, \dots, \mathcal{F}_r \subset \mathcal{P}([n])$  are intersecting families, and  $0 < p \leq 1/2$ , then*

$$\mu_p(\mathcal{F}_1 \cup \dots \cup \mathcal{F}_r) \leq 1 - (1 - p)^r.$$

*Proof.* By replacing  $\mathcal{F}_i$  with  $\mathcal{F}_i^\uparrow$  for each  $i$ , if necessary, we may assume that each  $\mathcal{F}_i$  is increasing. For each  $i$ , since  $\mathcal{F}_i$  is intersecting, Theorem 6 implies that  $\mu_p(\mathcal{F}_i) \leq p$ , and therefore  $\mu_p(\mathcal{F}_i^c) \geq 1 - p$ . Hence, using Corollary 8 (applied to the down-sets  $\mathcal{F}_1^c, \dots, \mathcal{F}_r^c$ ), we have

$$\mu_p(\mathcal{F}_1 \cup \dots \cup \mathcal{F}_r) = 1 - \mu_p(\mathcal{F}_1^c \cap \dots \cap \mathcal{F}_r^c) \leq 1 - \prod_{i=1}^r \mu_p(\mathcal{F}_i^c) \leq 1 - (1 - p)^r,$$

as required.  $\square$

Clearly, Lemma 19 is sharp, as can be seen by taking  $\mathcal{F}_i = \{S \subset [n] : i \in S\}$  for each  $i \in [r]$ .

**Upper bounds on linear combinations of sizes of cross-intersecting families.**

**Lemma 20.** *For each constant  $C_1 > 0$ , there exists a constant  $C_2 = C_2(C_1) > 0$  such that the following holds. Let  $\frac{n}{C_1} < k_1 < \frac{n}{2} - C_2$ ,  $\frac{n}{C_1} < k_2 < \frac{n}{2} - C_2$  with  $|k_1 - k_2| \leq C_1$ , and let  $t_0 \geq C_2 / \log\left(\frac{n-k_1}{k_1}\right)$ . Suppose that  $\mathcal{G}_1 \subset \binom{[n]}{k_1}$ ,  $\mathcal{G}_2 \subset \binom{[n]}{k_2}$  are cross-intersecting families with  $|\mathcal{G}_1| \leq \binom{n-t_0}{k_1-t_0}$ . Then*

$$|\mathcal{G}_2| + C_1 |\mathcal{G}_1| \leq \binom{n}{k_2},$$

*and equality holds only if  $\mathcal{G}_1 = \emptyset$ .*

*Proof.* Choose  $t \in \mathbb{N}$  such that  $\binom{n-t-1}{k_1-t-1} \leq |\mathcal{G}_1| \leq \binom{n-t}{k_1-t}$ . Note that  $t \geq t_0 \geq C_2 / \log\left(\frac{n-k_1}{k_1}\right)$ . By Lemma 15, we have  $|\mathcal{G}_2| \leq \binom{n}{k_2} - \binom{n-t-1}{k_2}$ . So it suffices to prove that  $\frac{\binom{n-t-1}{k_2}}{\binom{n-t}{k_1-t}} > C_1$ .



Observe that

$$\frac{\binom{n-t-1}{k_2}}{\binom{n-t}{k_1-t}} = \Theta_{C_1} \left( \frac{\binom{n-t}{k_1}}{\binom{n-t}{k_1-t}} \right),$$

and

$$\frac{\binom{n-t}{k_1}}{\binom{n-t}{k_1-t}} = \frac{(n-k_1) \cdot (n-k_1-1) \cdot \dots \cdot (n-k_1-t+1)}{(k_1) \cdot (k_1-1) \cdot \dots \cdot (k_1-t+1)} \geq \left( \frac{n-k_1}{k_1} \right)^t \geq 2^{C_2}.$$

Hence,

$$\frac{\binom{n-t-1}{k_2}}{\binom{n-t}{k_1-t}} = \Theta_{C_1} \left( \frac{\binom{n-t}{k_1}}{\binom{n-t}{k_1-t}} \right) = \Omega_{C_1}(2^{C_2}) > C_1,$$

provided  $C_2$  is sufficiently large depending on  $C_1$ , as required.  $\square$

**Approximate containment in dictatorships.** We now show that if  $\mathcal{F} = \mathcal{F}_1 \cup \dots \cup \mathcal{F}_r$  with  $\mathcal{F}_i \subset \binom{[n]}{k}$  an intersecting family for each  $i \in [r]$ , and  $|\mathcal{F}| \approx \binom{n}{k} - \binom{n-r}{k}$ , then not only is  $\mathcal{F}$  well-approximated by  $(\text{OR}_R)^{(k)}$  for some  $R \in \binom{[n]}{r}$ , but in fact each  $\mathcal{F}_i$  is well-approximated by a (different) dictatorship  $\mathcal{D}_j$  (with  $j \in R$ ). Specifically, we prove the following.

**Lemma 21.** *Let  $s \geq C_0 \sqrt{\log k}$  for some sufficiently large absolute constant  $C_0$ , let  $n \geq 2k + s\sqrt{k}$ , and let  $\mathcal{F} = \mathcal{F}_1 \cup \dots \cup \mathcal{F}_r$ , where  $\mathcal{F}_i \subset \binom{[n]}{k}$  is an intersecting family for each  $i \in [r]$ . If  $|\mathcal{F}| \geq \binom{n}{k} - \binom{n-r}{k} - \binom{n-r-s^2}{k-1}$ , then there exists a set  $R \in \binom{[n]}{r}$  and a permutation  $\pi \in \text{Sym}(R)$  such that  $|\mathcal{F}_i^{\emptyset}_{\{\pi(i)\}}| \leq 2^{2r} e^{-\Theta(s^2)} \binom{n-1}{k}$  for each  $i \in R$ .*

*Proof.* First note that by Theorem 5, we have  $|\mathcal{F}_R^{\emptyset}| \leq 2^r e^{-\Theta(s^2)} \binom{n-r}{k}$  for some  $R \in \binom{[n]}{r}$ ; without loss of generality, we may assume that  $R = [r]$ . Hence,

$$\begin{aligned} |\mathcal{F}_{[r]}^{\{j\}}| &\geq \binom{n-r}{k-1} \left( 1 - 2^r e^{-\Theta(s^2)} \frac{n-r-k+1}{k} \right) \\ &= \binom{n-r}{k-1} \left( 1 - 2^r e^{-\Theta(s^2)} \right) \end{aligned}$$

for each  $j \in [r]$ .

Note that for each  $j_1 \neq j_2 \in [r]$ , the families  $(\mathcal{F}_i)_{[r]}^{\{j_1\}}, (\mathcal{F}_i)_{[r]}^{\{j_2\}}$  are cross-intersecting. So we may assume, without loss of generality, that  $\mu_{\frac{1}{2}} \left( \left( (\mathcal{F}_i)_{[r]}^{\{j\}} \right)^{\uparrow} \right) \leq \frac{1}{2}$  for any  $j \neq i$ .

Fix  $j \in [r]$ . By Lemma 14 together with the Chernoff bound (2.2), we have  $\mu_{\frac{1}{2}} \left( \left( (\mathcal{F}_{[r]}^{\{j\}} \right)^{\uparrow} \right) \geq 1 - 2^r e^{-\Theta(s^2)}$ . Using Corollary 8, we have

$$1 - \mu_{\frac{1}{2}} \left( \left( (\mathcal{F}_{[r]}^{\{j\}} \right)^{\uparrow} \right) \geq \prod_{i=1}^r \left( 1 - \mu_{\frac{1}{2}} \left( \left( (\mathcal{F}_i)_{[r]}^{\{j\}} \right)^{\uparrow} \right) \right) \geq \left( \frac{1}{2} \right)^{r-1} \left( 1 - \mu_{\frac{1}{2}} \left( \left( (\mathcal{F}_j)_{[r]}^{\{j\}} \right)^{\uparrow} \right) \right).$$

Rearranging, we obtain

$$\mu_{\frac{1}{2}} \left( \left( (\mathcal{F}_j)_{[r]}^{\{j\}} \right)^{\uparrow} \right) \geq \mu_{\frac{1}{2}} \left( \left( (\mathcal{F}_j)_{[r]}^{\{j\}} \right)^{\uparrow} \right) \geq 1 - 2^{2r} e^{-\Theta(s^2)}.$$

Hence  $\mu_{\frac{1}{2}} \left( \left( (\mathcal{F}_j)_{\{j\}}^\emptyset \right)^\uparrow \right) \leq 2^{2r} e^{-\Theta(s^2)}$  and the lemma follows from Lemma 14 and the Chernoff bound (2.2).  $\square$

*Proof of Theorem 3.* Let  $\mathcal{F} = \mathcal{F}_1 \cup \dots \cup \mathcal{F}_i$ , where  $\mathcal{F}_i \subset \binom{[n]}{k}$  is an intersecting family for each  $i \in [r]$ , and suppose that  $|\mathcal{F}| \geq \binom{n}{k} - \binom{n-r}{k}$ . Then  $\mathcal{F}$  cannot contain  $r+1$  pairwise disjoint sets, so by Theorem 4, if  $n \geq (2r+1)k - r$ , we have  $|\mathcal{F}| \leq \binom{n}{k} - \binom{n-r}{k}$ , with equality only if  $\mathcal{F} = (\text{OR}_R)^{(k)}$  for some  $R \in \binom{[n]}{r}$ . Hence, we may assume throughout that  $n \leq (2r+1)k - r - 1$ . Moreover, by choosing  $C = C(r)$  to be sufficiently large, we may assume throughout that  $n \geq n_0(r)$  for any  $n_0(r) \in \mathbb{N}$ .

By Theorem 5 (applied with  $s = C(r)k^{1/6}$ , where  $C(r)$  is to be chosen later), Lemma 19 and Lemma 21, there exists a set  $R \in \binom{[n]}{r}$  and a permutation  $\pi \in \text{Sym}(R)$  such that

$$\left| (\mathcal{F}_i)_{\{\pi(i)\}}^\emptyset \right| \leq 2^{2r} e^{-\Theta(s^2)} \binom{n-1}{k}$$

for each  $i \in R$ . Without loss of generality, we may assume that  $R = [r]$  and  $\pi = \text{Id}$ , so that

$$\left| (\mathcal{F}_i)_{\{i\}}^\emptyset \right| \leq 2^{2r} e^{-\Theta(s^2)} \binom{n-1}{k}$$

for all  $i \in [r]$ .

Note that

$$|\mathcal{F}| \leq \binom{n}{k} - \binom{n-r}{k} + \sum_{j=1}^r \left| \left( (\mathcal{F}_j)_{\{j\}}^\emptyset \right) - \left( (\mathcal{F}_j)_{[r]}^{\{j\}} \right)^c \right|,$$

where  $\left( (\mathcal{F}_j)_{[r]}^{\{j\}} \right)^c := \binom{[n] \setminus [r]}{k-1} \setminus (\mathcal{F}_j)_{[r]}^{\{j\}}$ . Since

$$s^2 = \Omega_r \left( \frac{\sqrt{k}}{s} \right) = \Omega_r \left( 1 / \log \left( \frac{n-r-k}{k} \right) \right),$$

provided  $C = C(r)$  is sufficiently large we have

$$\left| (\mathcal{F}_j)_{[r]}^T \right| \leq 2^{2r} e^{-\Theta(s^2)} \binom{n-1}{k} \leq \binom{n-r-t_0}{k-|T|-t_0}$$

for all  $T \subset [r] \setminus \{j\}$  and all  $j \in [r]$ , where

$$t_0 \geq C_2 (\max\{2^{r-1}, 2r+1\}) / \log \left( \frac{n-r-k+|T|}{k-|T|} \right).$$

Hence, we may apply Lemma 20 with  $n-r$  in place of  $n$ ,  $C_1 = \max\{2^{r-1}, 2r+1\}$ ,  $k-r+1 \leq k_1 \leq k$  and  $k_2 = k-1$  to conclude that

$$(3.2) \quad \left| (\mathcal{F}_j)_{\{j\}}^\emptyset \right| - \left| \left( (\mathcal{F}_j)_{[r]}^{\{j\}} \right)^c \right| \leq 2^{r-1} \max_{T \subset [r] \setminus \{j\}} \left| (\mathcal{F}_j)_{[r]}^T \right| - \left| \left( (\mathcal{F}_j)_{[r]}^{\{j\}} \right)^c \right| \leq 0$$

for each  $j \in [r]$ . Hence,  $|\mathcal{F}| = \binom{n}{k} - \binom{n-r}{k}$ , so equality holds in (3.2) for each  $j \in [r]$ . Therefore, by Lemma 20,  $(\mathcal{F}_j)_{[r]}^T = \emptyset$  for all  $T \subset [r] \setminus \{j\}$ , i.e.  $(\mathcal{F}_j)_{\{j\}} = \emptyset$ , so  $\mathcal{F}_j \subset \mathcal{D}_j$  for all  $j \in [r]$ . Hence,  $\mathcal{F} \subset \text{OR}_{[r]}^{(k)}$ , so  $\mathcal{F} = (\text{OR}_{[r]}^{(k)})$ , as required.  $\square$

**Acknowledgements.** We would like to thank Nicholas Day for informing us of the open problem considered here.

## REFERENCES

- [1] R. Ahlswede and G. O. Katona, Contributions to the geometry of Hamming spaces, *Disc. Math.* **17** (1977), 1–22.
- [2] B. Bollobás, D. E. Daykin and P. Erdős, Sets of independent edges of a hypergraph, *Quart. J. Math. Oxford Ser. 2* **27** (1976), 25–32.
- [3] D. Ellis, N. Keller and N. Lifshitz, Stability versions of Erdős-Ko-Rado type problems, via isoperimetry, preprint. arXiv:1604.02160.
- [4] D. Ellis, N. Keller and N. Lifshitz, Stability versions of isoperimetric inequalities on the hypercube, in preparation.
- [5] P. Erdős, A problem on independent  $r$ -tuples, *Ann. Univ. Sci. Budapest* **8** (1965), 93–95.
- [6] P. Erdős, Problems and results on finite and infinite combinatorial analysis, in *Infinite and finite sets*, (A. Hajnal, R. Rado and V. T. Sós, Eds.), Proc. Colloq. Math. Soc. J. Bolyai 10 (Keszthely, Hungary, 1973), pp. 403–424, Bolyai–North-Holland, Amsterdam, 1976.
- [7] P. Erdős, C. Ko and R. Rado, Intersection theorems for systems of finite sets, *Quart. J. Math. Oxford, Ser. 2* **12** (1961), 313–320.
- [8] C. M. Fortuin, P. W. Kasteleyn and J. Ginibre, Correlation inequalities on some partially ordered sets, *Commun. Math. Phys.* **22** (1971), 89–103.
- [9] P. Frankl, Erdős-Ko-Rado theorem with conditions on the maximal degree, *J. Combin. Theory, Ser. A* **46** (1987), 252–263.
- [10] P. Frankl, Improved bounds for Erdős’ Matching Conjecture, *J. Combin. Theory, Ser. A* **120** (2013), 1068–1072.
- [11] P. Frankl and Z. Füredi, Extremal Problems concerning Kneser Graphs, *J. Combin. Theory, Ser. B* **40** (1986), 270–285.
- [12] T. E. Harris, A lower bound for the critical probability in a certain percolation, *Math. Proc. Camb. Phil. Soc.* **56** (1960), 13–20.
- [13] H. Huang, P. Loh and B. Sudakov, The size of a hypergraph and its matching number, *Combin. Probab. Comput.* **21** (2012), 442–450.
- [14] J. Kahn and G. Kalai, Thresholds and expectation thresholds, *Combin. Probab. Comput.* **16** (2007), 495–502.
- [15] L. Russo, An approximate zero-one law, *Z. Wahrsch. Verw. Gebiete* **61** (1982), 129–139.

SCHOOL OF MATHEMATICAL SCIENCES, QUEEN MARY, UNIVERSITY OF LONDON, MILE END ROAD, LONDON, E1 4NS, UK.

*E-mail address:* d.ellis@qmul.ac.uk

DEPARTMENT OF MATHEMATICS, BAR ILAN UNIVERSITY, RAMAT GAN, ISRAEL.

*E-mail address:* noamlifshitz@gmail.com